American Journal of Medical Genetics 64:424–427 (1996)

Trinucleotide Repeats at the FRAXF Locus: Frequency and Distribution in the General Population

Jeanette J.A. Holden, Melissa Walker, Maryse Chalifoux, and Bradley N. White

Department of Psychiatry, Queen's University (J.J.A.H., M.W.) and Cytogenetics and DNA Research Laboratory, Ongwanada Resource Centre (J.J.A.H., M.C.), Kingston, Ontario; and Department of Biology (B.N.W.), McMaster University, Hamilton, Ontario, Canada

FRAXF, the third X-chromosomal fragile site to be cloned, has been shown to harbour a polymorphic compound triplet array: $(GCCGTC)_n$ $(GCC)_n$. Expansion and methylation of the GCC-repeat and the neighbouring CpG-rich region result in chromosomal fragility. DNAs from 500 anonymous consecutive newborn males were examined to determine the incidence of various repeat numbers. The range of repeats was from 10–38, with the most common alleles having 14 (52.7%), 12 (16.6%), 21 (9.0%), and 22 (5.2%) triplets. Based on the distribution of repeat numbers, we suggest that the 21-repeat allele resulted from hairpin formation involving 7 GCC-repeats in a 14-repeat allele, accompanied by polymerase slippage. Examination of dinucleotide repeats near the FRAXF repeat will be important in testing this hypothesis. Since the clinical phenotype, if any, of FRAXF is unknown, this database will also be valuable for comparisons with repeat numbers in individuals from special populations.

KEY WORDS: FRAXF, trinucleotide repeat expansion, fragile site, population study

INTRODUCTION

Fragile sites are defined as chromosomal regions which show an increased frequency of gaps and breaks resulting from specific tissue culture conditions [Sutherland, 1979]. The molecular basis for chromosome fragility has been greatly speculated upon since Sutherland's [1979] discovery of the role of tissue culture conditions on their expression, but the greatest breakthrough came in 1991 with the discovery of CGG-repeat expansion as the cause of the fragile X syndrome and the recognition that chromosomal fragility was associated with hypermethylation of CpGs both within the repeat and of the neighbouring CpG island [Verkerk et al., 1991; Fu et al., 1991; Kremer et al., 1991; Oberlé et al., 1991]. Since 1991, four other folate-sensitive fragile sites, including two X-chromosomal [Knight et al., 1993; Parrish et al., 1994; Ritchie et al., 1994] and two autosomal [Nancarrow et al., 1994, 1995; Jones et al., 1995] sites, have been characterized, all of which have CGG/GCC-repeat expansion and hypermethylation. The normal alleles at four of these fragile sites (FRAXA, FRAXE, FRAXF, and FRA16A) are highly polymorphic, having varying numbers of repeats, whereas at FRA11B the majority of alleles have the same number of repeats [Jones et al., 1995].

There are several issues still being addressed about CGG/GCC-repeat expansion and instability and the role (if any) of these sequences in gene expression. The identification of CGG/GCC-repeat binding proteins which bind either the methylated or the unmethylated forms [Richards et al., 1993] and the fact that at least the FMR1 CGG-repeat is conserved in mammals [Deelen et al., 1994] suggest that, in some cases, CGG/GCC-repeats may affect gene expression. King [1994] has suggested that variation in repeat numbers may reflect a normally advantageous regulatory system. He argues from a population and evolutionary standpoint that if these triplet repeat lengths somehow regulate the quantitative expression of the respective genes, then changes in the numbers of repeats would provide phenotypic variability with each generation. Determination of accurate frequencies and the distribution of different repeat number alleles for the different CGG/GCC-repeat containing fragile-site loci may assist in developing and testing models for the role of these sequences in gene expression. This information may also help in modeling how repeat-number variation oc-

Received for publication September 25, 1995; revision received December 28, 1995.

Address reprint requests to Jeanette J.A. Holden, Ph.D., Cytogenetics and DNA Research Laboratory, Ongwanada Resource Centre, 191 Portsmouth Avenue, Kingston, Ontario, Canada K7M 8A6.

curs and how it is usually maintained within a specific range of repeat sizes until a pathological threshold occurs.

To this end, we have examined 500 consecutive newborn males, studied anonymously, for the number of triplet repeats at the FRAXF locus. Based on the frequencies of the different alleles, we suggest that hairpin formation and polymerase slippage within one of these alleles led to much of the variability seen in apparently normal FRAXF triplet repeat alleles.

MATERIALS AND METHODS

Samples

Samples for this study consisted of anonymous consecutive newborn Guthrie spots from males obtained from the Ontario Ministry of Health, following ethics review at Queen's University. These samples have previously been used to determine the number of CGG-repeats at the FMR1 locus [Holden et al., 1995]. Prior to their initial use, the Guthrie cards with blood spots were autoclaved and stored dry until use [Holden et al., 1996a].

PCR Analyses

The primers used were as previously described [Parrish et al., 1994]. A hole punch was used to remove a small section (0.5 cm diameter) of the Guthrie cards, and these were placed in 100 μl of 1X PCR buffer, boiled for 30 minutes and stored at −20°C until use. Immediately prior to removing an aliquot for amplification, the samples were again placed in a boiling water bath for 10–15 minutes. Amplification was carried out using 2 μl of template in a 10 μl reaction containing 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 200 μM dATP, 200 μM dTTP, 200 μM dCTP, 50 μM dGTP, 150 mM 7-deaza-dGTP, 1.5 μCi (alpha-^{32}P)dCTP, 10% DMSO, 0.5 U Taq polymerase (BRL), and 5 pmoles of each of the primers. Each reaction was overlaid with mineral oil and denatured at 94°C for 5 minutes followed by 30 cycles of 94°C, 1.5 minutes; 65°C, 1 minute; 72°C, 2 minute; and 10 minutes at 75°C. Aliquots of each reaction were mixed with formamide loading buffer and heated at 95°C for 5 minutes and the products were separated by electrophoresis through a 4% denaturing polyacrylamide gel. Alleles were visualized by exposure to Kodak XAR-5 film.

Triplet repeat numbers were as calculated by Ritchie et al. [1994], who found three or four copies of a GC-CGTC array adjacent to the GCC-repeat region described by Parrish et al. [1994]. Thus, for example, repeat numbers of 14 could represent either of the following configurations: $(GCCGTC)_3$ $(GCC)_8$ or $(GCCGTC)_4$ $(GCC)_6$. As Parrish et al. [1994] did not observe variation in the number of GCCGTC arrays among those sequenced, they concluded that all variation resulted from the number of GCC-repeats. For comparisons of our results with those of Parrish et al. [1994], it is necessary to add six to their triplet repeat numbers.

RESULTS

Figure 1A, B shows typical results for FRAXF triplet repeat amplification from newborn blood spots. Amplification of the FRAXF triplet repeat region was achieved for all but one of the 500 consecutive samples tested. The range of repeats was 10–38, with two major peaks at 14 repeats (52.7%) and 12 repeats (16.6%) (Table I). A single individual had 13 repeats. The next most common alleles consisted of 21 (9.0%) and 22 (5.2%) repeats.

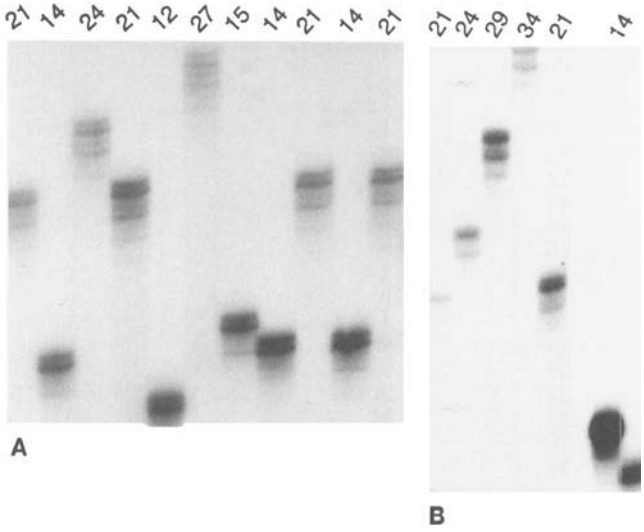

Fig. 1. **A**: A typical result obtained for the amplification of the triplet repeat region of the FRAXF locus from Guthrie spots from newborn males. **B**: Amplification products from large normal alleles (29 and 34 repeats) illustrating the presence of additional stutter or shadow bands compared to smaller normal alleles. The numbers at the top of the tracks indicate the number of triplet repeats in the sample.

Six individuals (1.2%) had an allele ≥30 repeats. We noticed that the majority of alleles greater than 28 repeats showed more shadow bands than those samples with 12 or 14 repeats (Fig. 1B), indicating possible mitotic instability.

Three individuals were found to have two distinct FRAXF triplet repeat alleles, both in the original test and using a second hole punch from the Guthrie cards. These were tested for the presence of ZFY Y-chromosome sequences [Witt and Erickson, 1989] and two were positive (data not presented). The other sample was tested at two additional Y-chromosomal sites and was negative.

DISCUSSION

We have successfully amplified the FRAXF triplet repeat from 499/500 blood spots taken from consecutive newborn boys, providing a frequency and distribution of repeat numbers in a general population. The most common allele of 14 repeats was found in 53% of the individuals, a somewhat higher proportion than the 42% reported by Parrish et al. [1994] in their study of 123 alleles from the CEPH families. The Canadian sample is ethnically diverse and it will therefore be of interest to examine the repeat in different ethnic groups, as this may provide information about the origin of the two major and other variant alleles.

Samples from three individuals showed two distinct FRAXF alleles. Two of these were shown to have Y-chromosome sequences, one of which also showed heterozygosity at both FMR1 and DXS102. The most likely

TABLE I. Distribution and Frequency of Triplet Repeats at the FRAXF Locus

Repeat number	N	Repeat number	N
10	1	25	6
11	0	26	6
12	83	27	5
13	1	28	4
14	263	29	0
15	8	30	1
16	3	31	2
17	2	32	0
18	4	33	1
19	7	34	1
20	7	35	0
21	45	36	0
22	26	37	0
23	11	38	1
24	13		

interpretation is that these samples are from XXY males, indicating that the frequency of XXY individuals in the samples tested was 2/500 (0.4%). Although this incidence may seem high compared to reported incidences of about 1/1,000 males, our findings are based on only a small sample size and may be spurious. The remaining sample with two FRAXF alleles did not appear to have Y-chromosome sequences and may be from an XX male, a male with a small duplication of Xq28, or a female. Although every attempt is made to ensure accurate labelling of the anonymous blood spots (two persons check the sex indicated on each Guthrie card prior to removal of identifying information and randomization of samples [Holden et al. 1996a]), we cannot eliminate the possibility that this sample is from a normal female. We do believe that the results do not reflect contamination of the blood spot from, for example, prolonged storage in contact with another Guthrie card or blood spot, since we have tested this and even paperclipping blood spots with different FMR1 alleles together for prolonged periods does not lead to sufficient contamination to be detected using our standard PCR protocols.

Parrish et al. [1994] were unable to amplify the FRAXF triplet repeat region from 4/79 parents from CEPH families. They suggested that this may be due to point mutations in the sequences recognized by the primers, microdeletions at the locus, or low-quality DNA preparations. Since only 1/500 samples we tested failed to give a PCR product upon repeated testing, and this sample had normal FMR1 and FRAXE triplet repeat alleles, we believe that microdeletions and sequence variations of the FRAXF primer regions are not common. Further, we may have identified a "premutation" or full mutation at the FRAXF locus and we are currently trying to determine whether this is the case by additional testing of the FRAXF region.

To date, very few families have been identified with FRAXF triplet repeat expansion. In those described by Parrish et al. [1994] and Ritchie et al. [1994], individuals expressing fragility at FRAXF had large expansions, exceeding 300 repeats. In these studies, no individuals were identified with what might be called "premutation" alleles (i.e., between "normal" size alleles and "full mutations"), borrowing the terminology used for FMR1 CGG-repeat expansions. Premutations would be those alleles which are expanded, but which are not methylated and do not result in chromosomal fragile site expression. Parrish et al. [1994] did, however, describe a CEPH family in which there was instability of a possible "grey zone" or "intermediate" allele: a mother heterozygous for 14 and 35 (termed 8 and 29 in their paper) repeat alleles, had five children who inherited the 35 repeat allele and one son who had both a 35 and a 36 repeat alleles, suggesting that alleles in this range can be unstable. The largest allele found in the present study had 38 repeats, and 5 other individuals had greater than 30 repeats. The X-rays for alleles $\geq$29 repeats frequently showed increased numbers of stutter bands, similar to what we often see for grey zone FMR1 alleles, particularly those with large pure CGG-tracts (unpublished observations). Since we are unable to do family studies on our samples, we cannot confirm stability or instability of these FRAXF alleles. However, it will be important to identify families with these larger alleles and to do family studies to examine their stability, as this may provide a clue as to the origin of variability within the FRAXF triplet repeat.

The FRAXF families examined by Parrish et al. [1994] and Ritchie et al. [1994] had both mentally handicapped and mentally normal individuals who expressed fragility at FRAXF and who had expanded FRAXF triplet repeats, and they concluded that expansion at the FRAXF locus does not result in mental impairment. The high incidence of 12- and 14-repeat alleles with only 1/501 alleles having 13 repeats suggests either an ancient mutational event or a possible important function for the FRAXF triplet repeat region. It is therefore important to do additional clinical assessments of the apparently normal individuals who are FRAXF-positive and to identify additional families with FRAXF triplet repeat expansion, since some clinical phenotypes (such as cancer susceptibility, psychiatric illness, attention deficit, etc.) may not have been recorded or recognized in these individuals. Assessment of neighbouring microsatellite loci is needed to establish whether the 12- and 14-repeat alleles arose recently or have an ancient origin.

Finally, there is the question of the mechanism of generating and maintaining FRAXF CGG/GCC repeat variability in the "normal" range of perhaps 10–30 repeats. Unlike the FMR1 repeat, which has a peak repeat number of 30, with smaller peaks of 20 and 40 repeats and interrupting AGGs at approximately 10, 20, and 30 repeats which are thought to stabilize the repeat [Eichler et al., 1994]; unlike the FRAXE [Knight et al., 1993; Holden et al., 1996b] and FRA16A [Nancarrow et al., 1995] repeats which are pure GCC-tracts showing a more "normal" distribution; and unlike the FRA11B repeat in which the majority of alleles have a single repeat number [Jones et al., 1995]; the distribution of FRAXF repeats shows two discreet sharp peaks at 12 and 14 repeats, with a broad peak at 21–24 repeats. Chen et al. [1995] found that both the C-rich and G-rich single strands of CGG/GCC repeats

form hairpins under physiological conditions and that C-rich strands are more likely to form these hairpin or slippage structures during replication, resulting in asymmetric repeat expansion. They also found that odd numbers of repeats tend to form higher proportions of hairpin structures than duplexes compared to the next higher even numbers of repeats. These findings suggest a possible model for the origin of the 21-repeat alleles at the FRAXF locus. We propose that the 21-repeat alleles originated from a 14-repeat allele [$(GCCGTC)_3(GCC)_8$], in which seven of the eight GCCs formed a hairpin or slippage structure during replication. Rereplication of the seven GCC repeats would lead to a total length of 15 GCC triplets, or a triplet repeat length of 21-repeats. Analysis of flanking microsatellite loci and the identification of founder chromosomes for the 21–24 repeat alleles would support this hypothesis.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Ontario Mental Health Foundation to J.J.A.H. and B.N.W.

REFERENCES

Chen X, Mariappan SVS, Catasti P, Ratliff R, Moyzis RK, Laayoun A, Smith SS, Bradbury EM, Gupta G (1995): Hairpins are formed by the single DNA strands of the fragile X repeats: Structure and biological implications. Proc Natl Acad Sci USA 92:5199–5203.

Deelen W, Bakker C, Halley DJJ, Oostra DA (1994): Conservation of CGG region in FMR1 gene in mammals. Am J Med Genet 51: 513–516.

Eichler EE, Holden JJA, Popvich BW, Relss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994): Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Na Genet 8:88–94.

Fu YH, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJMH. Holden JJA, Fenwick RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991): Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. Cell 67:1047–1058.

Holden JJA, Chalifoux M, Wing M, Julien-Inalsingh C, Scott E, Fidler K, Swift I, Maidment B, White BN (1995): Distribution and frequency of FMR1 CGG repeat number in institutionalized developmentally disabled individuals. Devel Brain Dysfunction 8: 302–309.

Holden JJA, Chalifoux M, Wing M, Julien-Inalsingh C, White BN (1996a): A rapid, reliable and inexpensive method for detection of di- and trinucleotide repeat marker and disease loci from dried blood spots. Am J Med Genet 64:313–318.

Holden JJA, Julien-Inalsingh C, Chalifoux M, Wing M, Scott E, Fidler K, Swift I, Maidment B, Knight SJL, Davies KE, White BN (1996b): Trinucleotide repeat expansion in the FRAXE locus is not common among institutionalized individuals with non-specific developmental disabilities. Am J Med Genet 64:420–423.

Jones C, Penny L, Mattina T, Yu S, Bakr E, Voullaire L, Langdon WY, Sutherland GR, Richards RI, Tunnacliffe A (1995): Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. Nature 376:145–149.

King DG (1994): Triplet repeat DNA as a highly mutable regulatory mechanism. Science 263:595–596.

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembry ME, Holland J, Oostra BA, Bobrow M, Davies KE (1993): Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74:127–134.

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991): Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711–1718.

Nancarrow JK, Kremer E, Holman K, Eyre H, Doggett NA, Le Paslier D, Callen DF, Sutherland GR, Richards RI (1994): Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. Science 264:1938–1941.

Nancarrow JK, Holman K, Mangelsdorf M, Hori T, Denton M, Sutherland GR, Richards RI (1995): Molecular basis of p(CCG) repeat instability at the FRA16A fragile site locus. Hum Mol Genet 4: 367–372.

Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, Bertheas MF, Mandel JL (1991): Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102.

Parrish JE, Oostra BA, Verkerk AJMH, Richards CS, Reynolds J, Spikes AS, Shaffer LG, Nelson DL (1994): Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. Nature Genet 8:229–235.

Richards RI, Holman K, Yu S, Sutherland GR (1993): Fragile X syndrome unstable element, $p(CCG)_n$, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. Hum Mol Genet 2:1429–1435.

Ritchie RJ, Knight SJL, Hirst MC, Grewal PK, Bobrow M, Cross GS, Davies KE (1994): The cloning of FRAXF: Trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. Hum Mol Genet 3:2115–2121.

Sutherland GR (1979): Heritable fragile sites on human chromosomes. I. Factors affecting expression in lymphocyte culture. Am J Hum Genet 31:125–135.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DPA, Pizuti A, Reiner HA, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen GJB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST (1991): Identification of a gene (FMR-1) containing a CGG-repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914.

Witt M, Erickson RP (1989): A rapid method for detection of Y-chromosomal DNA from dried blood specimens by the polymerase chain reaction. Hum Genet 82:271–274.